The Whereabouts Diary

Gabriella Castelli, Marco Mamei, Alberto Rosi

Dipartimento di Scienze e Metodi dell'Ingegneria – Università di Modena e Reggio Emilia
Via Amendola 2 – 42100 Reggio Emilia – ITALY
{gabriella.castelli, marco.mamei, alberto.rosi}@unimore.it

Abstract. The user profile is one of the main context-information in a wide range of pervasive computing applications. Modern handheld devices provided with localization capabilities could automatically create a diary of user's whereabouts and use that information as a surrogate (or a complement) of the user profile. The places we go, in fact, reveal also something about us, for example, two persons can be matched as compatible given the fact they visit the same places. Web-retrieved information, and the temporal patterns with which different places are visited, can be used to automatically define meaningful semantic labels to the visited places. In our work we used geocoding and white-pages Web-services to extract information about a place, and Bayesian networks to classify places on the basis of the time in which they have been visited. In this paper we describe the general idea at the basis of the whereabouts diary, discuss our implementation, and present experimental results. Finally, several applications that can exploit the diary are illustrated.

1 Introduction

The recent diffusion of handheld devices and smart phones equipped with localization capabilities[1] is opening new scenarios in the development of context-aware services. The location itself is an extremely useful source of information: location-based services and location-based information retrieval allow to get resources that are relevant and practically accessible given the actual location of the user [BenM03, CasR07]. More than that, the places we go can reveal also something about us, and can be used as a surrogate or a complement to form a better user profile. For example, a matchmaking application could infer that two persons are compatible given the fact that they visit almost the same places. Moreover, if the places are tagged semantically (e.g., work, home, pub, etc.) the application could infer more advanced relationships among the persons. For example, two persons visiting the same "work" place could be marked as colleagues, while persons visiting the same "home" place could be marked as relatives.

[1] The U.S. E911 and European E112 initiatives require localization capability for calls placed to emergency services by mobile phones.

In this paper we present the ideas and a first prototype implementation of the *whereabouts diary*: an application, running on a GPS-equipped handheld device that records the list of relevant places visited by the user. The diary runs autonomously without requiring user's interactions and is able to classify *semantically* the places being visited in an unsupervised way. Relevant places can be extracted by considering clusters and dropouts in the GPS signal (typically indicating the user staying in a place or entering in a building). Semantic information can be added by exploiting the structure of people daily routine. For example, the place where the user usually spends night-time can be tagged semantically as "home", while the place where the user usually goes from 8am to 6pm can be tagged as "work". Specifically, we realized a set of Bayesian networks to diagnose the kind of place given the temporal pattern of user visits. Further information can be extracted by geocoding the place and mining the Web in search for relevant information. For example, the fact that the user was at the coordinates (lon: -80.239, lat: 25.955) on the 02/04/07 evening, can be easily geocoded to infer that the user was actually at the Dolphin Stadium in South Florida. Moreover, such information could be further refined extracting from the Web the fact that the SuperBowl was actually playing that night in that stadium. The result is a diary describing the user daily life and that could provide useful profile information to other applications.

The rest of this paper is organized as follows: Section 2 describes the general idea and our current implementation of the whereabouts diary. Section 3 presents some experiments we conducted to test the effectiveness of our prototype. Section 4 discusses some applications that could be realized with the diary. Section 5 presents some related work. Section 6 concludes and presents future work to improve the diary.

2 The Whereabouts Diary

In this section we first present the conceptual idea beneath the whereabouts diary, then we detail our current implementation.

2.1 General Idea

The construction of the whereabouts diary is an incremental process. Starting from the log of the GPS readings (or of other kind of localization devices), it is possible to run segmentation and clustering algorithms to infer the places where the user spends most of his time [HigC05]. The result of this first operation is a list of places described in terms of longitude and latitude, and a list of time intervals associated to each of the coordinates indicating when the user has been there. This first process creates a diary like the one presented in Fig. 1.

Longitude	Latitude	Time
-73.974	40.763	July, 4, 2006, 4:35pm-5:41pm
…	…	…

Fig. 1 Diary based on GPS coordinates.

A simple list of coordinates is only partially informative and the need of translating from positions to places (i.e., adding semantic meaningful tags to the discovered coordinates) has been widely recognized [Hig03]. A diary containing information like "*the user was at home*" rather than "*the user was at coordinates (10.873, 44.630)*" would be naturally much more informative and easy to use in context-aware applications.

A first step to in the process of adding semantic information would be to translate from coordinates to addresses. This can be done via standard tracking and geocoding services (as common GPS navigators do). However, because of errors in GPS localization and errors in the process of segmenting and clustering the GPS readings to identify relevant places, in most of the situations, it will not be possible to identify the unique address where the user is located, and only a partial estimate can be given (e.g., all the addresses within 10 meters from a given place are actually taken into consideration). This second step converts the diary in the one depicted in Fig. 2.

Place	**Time**
123, 5th Ave, NY, USA	July, 4, 2006, 4:35pm-5:41pm
4,5,…,21, 26th St., NY, USA	July 6, 2006, 7:00am – 8:00am
…	…

Fig. 2 Diary based on addresses. Because of GPS errors multiple addresses can be associated to a single place.

A third step can try to mine the Web to identify what is in a particular address. The primary source of information in this context would come from yellow- and white-pages services. However, due to the aforementioned localization errors, this process will return in some of the cases a list of all the businesses performed in the geocoded addresses. Still, in some situations a single exact match could be retrieved like in the case of the user being in a big stadium or entering a big shopping mall.

Even more semantic information could derive by searching relevant events that happened in that place at that time. For example, it could be possible to extract from the Web the fact that "the 4th of July parade" took place near the geocoded location at the same time the user was there. This process could create a diary like the one depicted in Fig. 3.

Place	**Time**
4th July Parade	July, 4, 2006, 4:35pm-5:41pm
126, 13th St., NY, USA	July 5, 2006, 11:00pm – 7:00am
Starbucks Coffee || Uno's Pizza	July 6, 2006, 7:00am – 8:00am
….	…

Fig. 3 Diary based on places. Because of errors in the previous phases, multiple businesses can be associated to a single place.

Finally, if the user activities are profiled in some way (e.g., the diary may know a priori that the user tends to stay at home at night), then the diary application can give labels to places by looking at the temporal patterns in which places are visited. For example, the place most visited at night during weekdays can be meaningfully labeled as "Home". Such kind of analysis can be also used in combination with commonsense information

[Liu04] to disambiguate between alternative retrieved places. For example, in the table in Fig. 3 the ambiguity among “Starbucks Coffee” and “Uno’s Pizza” can be resolved (at least from a probabilistic point of view) in favor of the former, in consideration of the fact the place has been visited from 7:00 am to 8:00 am.

Of course, should other kind of sensing devices be available (e.g., RFID and NFC – Near Field Communication – readers), classification could use such information to better identify the places. For example, a powerful source of data could come from credit card transaction records that would identify not only in which shop the user has been, but also what he has bought (some recent proposals in the context of mobile wallet applications go in this direction [Fit04]). In the end, the combination of all the above steps leads to a diary close to the one in Fig. 4.

Place	Time
4th July Parade	July, 4, 2006, 4:35pm-5:41pm
Home	July 5, 2006, 11pm – 7am
Starbucks Coffee	July 5, 2006, 7am – 8am
….	…

Fig. 4 Diary based on personalized places.

In its final form the diary represents a powerful source of context information allowing to extrapolate user’s habits, preferences and routine behavior.

2.2 Implementation

The whereabouts diary can be built incrementally and automatically on the basis of the user’s trace of locations. In our current implementation the trace is acquired only via a GPS, however other localization mechanisms such as WiFi or Cell tower triangulation can be used as well [HigC05].

The first step is to segment user’s GPS trace to find “relevant” places. Following an approach similar to [MarS00, SchR06], we tagged as relevant those places for which either one of the following conditions apply:

1. the GPS signal is lost for at least T seconds and it is re-acquired later on at a distance of less than L meters from where it was lost. This reflects the situation in which a user enters a building and leaves it after some time. Some empirical evaluations let us to set T = 20 minutes, L = 20 meters. The constraint on time is important to wash out GPS signal glitches, the constraint on space is useful to avoid those situation in which the GPS has been shut down and the user moves away.
2. The GPS readings over a time window of W seconds are clustered within a radius of R meters from each other. This reflects the situation in which the user stays for a long time in a place like a park or a square. Some empirical evaluations let us to set W = 20 minutes, R = 100 meters.

The list of relevant places is built online and incrementally. When a set of coordinates meets one of the above criteria, the system looks in the list of the already discovered places for one closer than L = 10 meters to the coordinates. If such a place does not exist, a new place is created and the time of visit is recorded. If the place exists, the place

coordinates are averaged with the new coordinates, and if enough time has passed since the previous visit (30 min), the time of the new visit is recorded. The output of this algorithm is represented with Google Earth in Fig. 5.

Fig. 5 Path and relevant places depicted in Google Earth.

The coordinates associated to places have been translated into addresses using a geocoding service. Most of the geocoding services available online (e.g., that provided by the Google Maps API) translate addresses into coordinates. Instead, the diary needs the reverse operation: from coordinates to addresses. We developed a "reverse" geocoding for our region, on the basis of maps available from a commercial navigator software. Once such maps are available, the approach is rather straightforward: the coordinates are mapped to the closer map entry (i.e., address) being available.

To take into account GPS and geocoding inaccuracies, and the errors introduced by the place retrieving process, the diary application tries to reverse geocode all the addresses within a radius of 10m from the place being segmented. Thus, the diary actually creates a list of candidate addresses where the user has been.

The next step is to translate addresses into businesses (i.e., shops, offices, etc.). The ideal result is to have in the diary entries like "Starbuck's coffee", rather than "234, Marlborough st.". To perform this operation we screen-scraped information coming from a widely used online white-pages service[2] in our region allowing to query for who is at a given address. This operation is trivially achieved using the tools provided by the htmlparser[3] software. In particular, each geocoded address belonging to a given place (as provided by the previous step) is looked up in the white-pages and the corresponding business is retrieved. The result of this process is a set of entries labeled with the

[2] www.paginebianche.it
[3] htmlparser.sourceforge.net

possible businesses found in that place. Fig. 6 shows the result of this process for a given place in Google Earth. This translation process in not completely accurate, since several addresses are not listed in the white-pages (mainly due to privacy constraints). Still, the fact that most public businesses (like shops, etc.) are listed, while several private houses are not, allows to prune out a lot of unlikely addresses being discovered by the previous step. Private spaces like "home" – that are likely not to be listed in the white-pages – can be derived from other kind of analysis described below.

Fig. 6 Extracted information related to a given place.

The final step in our implementation is more challenging. The diary tries to automatically extract semantic information to describe the relevant places from a personal point of view.

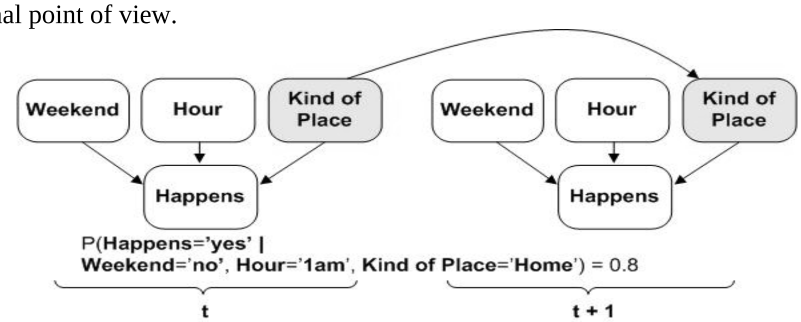

Fig. 7. Bayesian network to classify places. White nodes are those that will be provided as evidence.

To this end, for each place being identified in the first phase, the diary creates a Bayesian network to analyze the temporal pattern in which the place has been visited by the user (see Fig. 7).The Bayesian network is composed of 4 nodes.

1. The *weekend* node represents a boolean variable used to represent whether a given observation takes place in the weekend or not. This node is always observed on the basis of the information stored in the GPS signal. This information represents the variability in people behavior between weekdays and weekends.
2. The *hour* node is a 24-values discrete node storing the time of day. This node is always observed on the basis of the information stored in the GPS signal.
3. The *kind of place* node is a discrete node modeling what a given place is. In our implementation, we try to classify among 5 different kind of places: home, work, restaurant (to indicate any kind of dining place), pub (to indicate any kind of evening entertainment), and disco (to indicate any kind of late-night entertainment). This classification is rather arbitrary, and each user of the diary should provide the kinds of place that best match his habits. This node is never observed, and is inferred by probability computations.
4. The *happens* node is a boolean variable expressing whether the user visits that place at that time. This node is always observed on the basis of the outcomes of the diary localization phase.

The role of the Bayesian network is to encode the routine of the user daily life. This is done by compiling the probability distribution associated to the fact that the user, in a given moment, is in a certain kind of place. For example, the probability of the user being at home during weekdays is depicted in the table in Fig. 8.

Weekend = false, Kind of Place = home									
time	**11pm-6am**	**7am**	**8am**	**9am-1pm**	**2pm-5pm**	**6pm-7pm**	**8pm**	**9pm**	**10pm**
P(happens) = true	0.8	0.6	0.4	0.2	0.2	0.4	0.5	0.6	0.7

Fig. 8. Conditional probability table describing the probability of the user being at home during weekdays.

Similar tables can be created for other kind of places. In our current implementation, these tables are compiled by hand by the users that are asked to self-report the likelihood of being in a given kind of place at a given time. Such kind of data could be derived automatically also by a labeled trace of user's past whereabouts, using standard learning algorithms [PatL03]. Once the tables are filled in, basic inference operations in Bayesian networks will be used to derive the most likely kind of place given the visit pattern.

Specifically, when the diary previous phases identify that the user is visiting a place, the corresponding Bayesian network is retrieved, and the *weekend*, *hour*, *happens* nodes are set to their actual values (the *happens* node is trivially set to *true* to indicate that there is a visit). Then, the diary computes the probability distribution of the *kind of place* node. The newly computed distribution will be used as a prior for subsequent visits. This

naturally allows evidences to add up, actually enabling the Bayesian network to classify the places on the basis of the visit temporal pattern. The results of the Bayesian classifier for two places in our dataset is reported in Fig. 9.

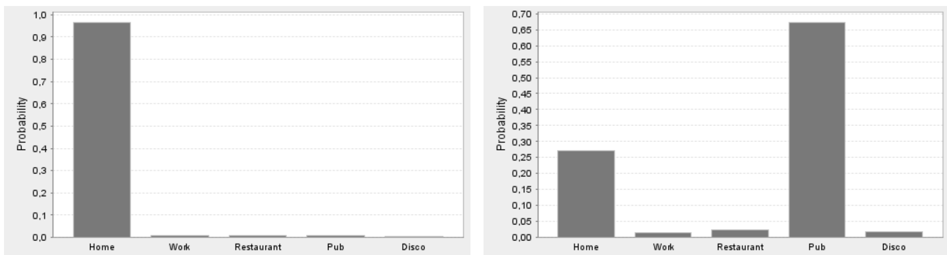

Fig. 9. Outputs of the Bayesian network classifier for two places.

3 Experiments

To test the effectiveness of the whereabouts diary, we collected GPS traces for three weeks from the three members of our research team (the authors) as they went about their normal lives. Each member carried either an i-mate PDA 2K smart phone, or a HP IPAQ RX3700 pda, connected with a Bluetooth GPS reader. GPS signal has been acquired at 0.1Hz and processed on the fly by the handheld device. Overall, we acquired about 90000 GPS poses amounting at 360 MB of data. In particular, the diary is a J2ME – personal profile application that:

1. Collects and stores the GPS trace log.
2. Runs the clustering algorithm to identify relevant places.
3. Creates a queue of the places to be resolved by the on-line (reverse) geocoding and white-pages services. Places are actually resolved whenever a WiFi connection becomes available.
4. Dynamically creates a Bayesian network for each newly discovered place.
5. When a relevant place is visited, the associated Bayesian network is retrieved, and the probability distribution to describe the kind of place is updated accordingly to the time of visit. Bayesian operations are implemented on the basis of software described in [MamN07].

During the data collection weeks, data collectors recorded ground-truth information about the places they have been.

All the data have been recorded in our region in Italy (none of the data collectors have been abroad during that time). The region is characterized by rather small and short buildings. On the one hand, this means that GPS signal problems related to urban canyons are less severe, and it is usually easy to get fairly accurate GPS positioning. On the other hand, buildings are packed closed to each others, and thus also small errors can produce wrong address recalls.

In the following, we present some results obtained by comparing the whereabouts diary entries, after some weeks of usage, with recorded ground-truth information.

In a first set of experiments, we tried to verify the accuracy of the algorithm to identify relevant places on the basis of the GPS trace log. Following an approach similar to [HigC05], we classify the incorrect results into: *(i) wrong*: the user is in a place, but the diary reports he is in a different place, *(ii) false negative*: the user is in a place, but the diary reports he is moving, *(iii) false positive*: the user is moving, but the diary reports he is in a place. The results of these first experiments are reported in Fig. 10, and they actually show the average of the results obtained by the data collectors. The results we obtained show that the algorithm is correct in 84.7% of the cases. This figure is coherent with the results presented in [HigC05] with regard to the A-S algorithm [AshS03] that is the one closer to our implementation. The high-percentage of false negative results (compared to the other cases) is mainly due to the fact sometimes the GPS takes a long time before acquiring the signal. Thus, it can happen that a user leaves a building, and the trace of the GPS is acquired only when he is already far away. In such a situation the place is not detected given the constraint on the maximum distance of spatial disconnection described in Sect. 2.2.

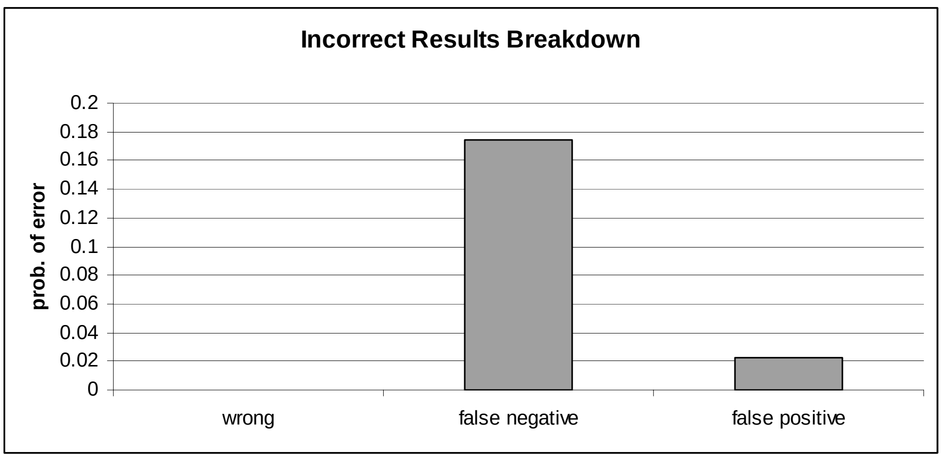

Fig. 10. Errors in the algorithm to identify relevant places on the basis of the GPS trace log. Place identification is correct in 84.7% of the cases. Errors are divided as reported in the graph.

In a second set of experiments, we tried to verify the results of the (reverse) geocoding service. Basically, the idea is to verify the impact of localization errors in the process of geocoding. It is worth noticing that the maps we used to perform this operation record only the first and the last number of a street segment and span, uniformly, all the other numbers among the segment. This of course introduces further errors in that it does not take in to account the differences in the sizes of the buildings.

Since the place discovery algorithm clusters together points that are closer than 10 meters, we counted the number of addresses retrieved within a circle of 10 meters radius centered at the relevant place. The results of this operations are displayed in Fig. 11, and highlight two aspects. On the one hand, the address of almost half of the places can be retrieved uniquely (this is the case of large buildings – like the departments of our university). On the other hand, some places produce more than 10 associated addresses. This is the case of small buildings in the center of the city. It is fair to report that these

distributions are rather preliminary since they are based on a dataset of only 25 places (those identified by the diaries of the 3 data collectors). We are currently conducting a more extensive data collection process that would allow us to identify more stable distributions.

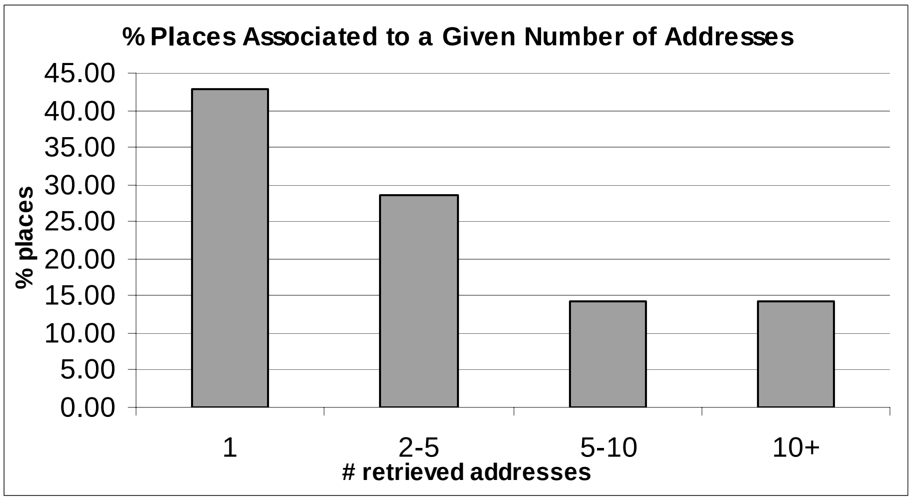

Fig. 11. Percent of relevant places corresponding to a given number of addresses.

In a third group of experiments, we tried to evaluate the performance of retrieving the businesses performed on a given place via the white-pages service. Comparing the results with the ground-truth annotations, we first tried to determine whether the correct place is retrieved. With disappointment, we verified that the actual place can be retrieved in only 40% of the cases. This is either due to localization or white-pages errors. Moreover, due to the multiplicity of addresses being discovered several businesses can be assigned to a given place. In Fig. 12, we report the distribution of the number of businesses found for a given place. It is easy to see that some addresses are not listed in the white-pages, since there are no places with more than 10 retrieved businesses. In addition, It is worth reporting that the number of businesses being retrieved is almost independent of whether the correct place has been found or not. Some places, in fact, return a long list of candidate entries not containing the correct one. The main source of errors of this phase is related to the white-pages interface and how it handles street numbers. For example, one puzzling error we found, involved the Juta bar in the center of our city. Although the place was acquired correctly, the address geocoded correctly, and the Juta bar is listed in the white-pages service, the diary was not able to extract such information. We found that the problem is about places having more than one street number (for example, the Juta bar is a long building located from 87 to 95 of Taglio st.). Unfortunately, the Web interface to the white-pages allows only to enter a single street number, and white-pages matching mechanism does not take into account the fact that e.g., 89 Taglio st. is within the range of Juta addresses, thus it produces no results. In our future work, we plan to solve these glitches by querying multiple white- and yellow-pages services and merging the results.

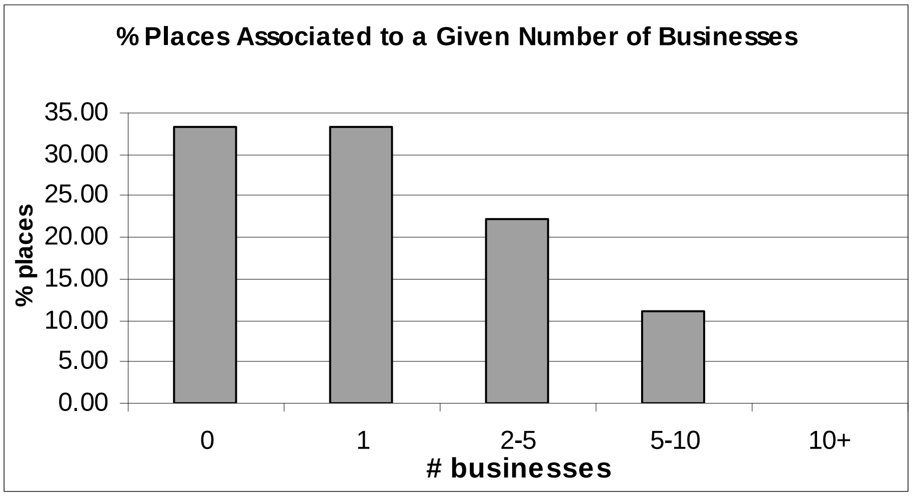

Fig. 12. Percent of relevant places corresponding to a given number of businesses being retrieved.

Finally, the last set of experiments verified the results of the Bayesian classification. Overall, our approach classifies the places correctly in 64% of the cases. In order to better analyze the results we tried to assess the confidence of the diary in its own classification – most probable estimate (MPE). To this end, we compute the information entropy of the resulting distributions. The lower the entropy, the more the system is confident about the MPE (i.e., the distribution peaks on the MPE value). More in detail, we separated the entropies related to the distributions that produced a correct MPE from those that produced a wrong MPE. For each of these two categories, we averaged together the entropies of the distributions producing the same MPE (see Fig. 13). Looking at the graph it is possible to notice that entropies related to wrong MPEs and higher than entropies related to correct MPEs. This is good and reflects the fact that the distributions associated to wrong estimates are less peaked, and thus the diary is less confident about its own classification. In fact, examining the wrong distributions in a lot of circumstances the distribution is bimodal: one peak is the correct one and another slightly more probable is incorrect. In such circumstances, it is likely that more observations on the temporal pattern of visits will correct the classification outcome.

Another interesting remark about the result in Fig. 13, relates to the average entropy associated to the different kind of places. Not surprisingly, *home* and *work* places have lower entropies since their associated temporal pattern of visits is defined more precisely. On the contrary, places like *pubs*, *restaurants* and *discos* have a more flexible pattern of visits and thus they are classified less precisely and tend to produce higher entropies.

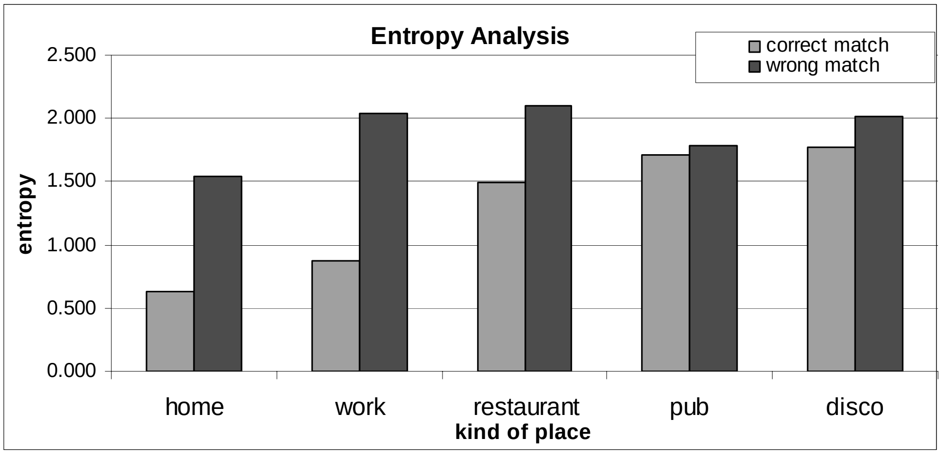

Fig. 13. Information entropy of the resulting distributions for the various kind of places. The information entropy for a 5-value discrete distribution ranges from 0 to 2.32 (flat distribution).

4 Applications

Once the diary is available, it can be used as a surrogate of the user profile in several applications.

Matchmaking. The automatic exchange of diaries between users can enforce powerful matchmaking capabilities. An application built on top of the whereabouts diary could try to match the places visited by the users to understand whether they are visiting similar places or not. The similarity between the places being visited can be a useful measure of the affinity between the two users. Similarly to what suggested in [KosN06], the diaries can be useful to in the process of building common ground for face to face interactions. For example, a user could be notified that the person next to her has been to the same concert one week ago. With this regard, it is important to remark that the use of diaries enriched with semantic information is really useful in this context. For example, the application can automatically discriminate among colleagues (people that spend time in the same work places), relatives (people that spend time in the same home places) and friends (people that spend time in the same pub and disco places). It would not be possible to conduct such kind of analysis on the basis of a purely geographic diary that only computes overlapping areas among people.

Places for tourists. Tourist scenarios have always motivated ubiquitous computing applications. These kind of applications being deeply related to user profile and context-awareness can naturally take advantage of the whereabouts diary. An application could match the places visited by the user across a database of tourist attraction and suggest possible next visits. For example, if the diary indicates that the user likes visiting museums, the application could suggest visit other museums of a city. In addition, the exchange of diaries among tourists could allow to exchange experiences, hints and recommendations. For example, backpackers traveling northward could fruitfully

exchange diaries with backpackers traveling southward to obtain information about where they had just been [AxuV06]. Also in this case, the extraction of semantic information (e.g., retrieving shops and other businesses in a given area) could provide a notable added value to the information being communicated.

Shop Assistant: A shop recommendation service could try to infer users preferences on the basis of shops visited by the user. The scenario for this application could be an open-air market, for example a typical flea-market, or a city shopping district. The system can know in advance the location and the typology of goods held by any expositor and shops (as possibly provided by a sort of yellow-pages service). An application running on the whereabouts diary can trace the user's movements along the first visited places, and compose a sort of user shopping profile. Starting from this profile, the system could try to predict future user movements and indicate him the way to expositors/shops that in its opinion better fulfill his wishes. The user profile could be continuously updated as user makes choices and moves across the area.

Pervasive Advertisement. Personalized advertisements offering users the most appropriate contents to suit their profile is a huge source of revenues for IT companies. The whereabouts diary could be a useful mechanism to apply personalized advertisement to ubiquitous computing scenarios. An application could show commercials to the user that are personalized on the basis of the diary. For example, if the user usually visits Au Bon Pain around noon, the application running on the user handheld device could present Au Bon Pain commercials at about 11am showing special offers, and indicating also those places that are in the neighborhood. In another setting, a wide screen display connected with a Bluetooth station could fetch the diaries of the people around and select the most appropriate commercials for the present population. In both these situations the whereabouts diary can provide useful information to compose the user profile. In addition, it is interesting to notice that the diary can provide useful feedback on whether the advertisement had been fruitful or not. For example, comparing the places visited by the user after seeing a commercial, it would be possible to infer whether the user followed the advertisement and actually went to Au Bon Pain.

Predict User Destination and Anomaly Detection. This application is based on the general assumption that people routinely performs repetitive actions (e.g., usually people go to work/school in the morning and come back in the evening, people have lunch around noon, etc.). On this basis, an application can learn from the whereabouts diary the user motion routine, and thus predict where the user will go next and eventually trigger alarms if anomalous deviations from the normal track are detected [PatL03, PatL04]. Such a service may be very useful for very young, elder or disabled people in order to support their independent living. For example, a child goes to school by himself; the application could monitor his movements alerting himself and his parents if notable deviations from the usual track are detected (the child may got lost or may have caught the wrong bus).

Whereabouts Routing. Some interesting application related to the prediction of user next movements are presented in [AshS03, RomH06]. Nodes connectivity and performance in mobile ad-hoc network scenarios could be notably enhanced by having nodes to communicate with each other their likely future movements. For example, an application could avoid initiating to exchange files with a node that is predicted to be out of the wireless range in few seconds. In another setting, this idea could be fruitfully exploited to enable network routing in the presence of large disconnections among

nodes. For example, a user ‘A’ willing to send a message to a user ‘B’, that is currently not reachable because of network partitions, could inspect the motion model of the user ‘B’ and retrieve the fact that ‘B’ at that time is usually in a given place. On this basis, the user ‘A’ could search for other reachable users that (according to their motion model) are going to the same place and relay the message to them. This approach can create an interesting routing mechanism that allows the messages to flow also in presence of large network partitions. Of course, the diary can be fruitfully applied to this context allowing the routing algorithm to predict where the user will go next based on users’ past history.

Personalized Navigation. MyRoute [PatC06] is an application to produce personalized navigation routes with the goal of reducing route complexity and cognitive burden. This is achieved by creating user specific routes on the basis of his previous knowledge about familiar routes and landmarks. In particular, MyRoute works by compressing directions coming from traditional navigation software into a single contextualized step (e.g., drive to work). The whereabouts diary could naturally support MyRoute by providing the information about users’ relevant places.

Context-aware Instant Messaging. The messenger application Nomatic [PatX06] proposes to exploit places’ semantic labels as meaningful status information in messenger applications. Such automatic labels could be used in combination of the standard (“on line”, “not at PC”, “busy” labels) to better express the actual status of the user. More in general, a wide range of ubiquitous computing applications can be enhanced by incorporating information about localization. For example, mobile phones can be automatically switched off inside a theatre, or could connect with the car speakers when the user is driving. Contact-management software could be programmed to automatically exchange business cards when the current place is recognized as work by both the users. Of course, the whereabouts diary could provide all these applications with suitable information.

Memory-aid. Memory-aid applications can take a notable advantage from the availability of the whereabouts diary. Specifically, notes, to-do-lists and reminders can be associated to specific places and be triggered by the user being in there. Moreover, the whereabouts diary can offer predictions on where the user is going next, so that reminders can be triggered also before the user reaches his actual destination. For example, if the diary predicts that the user will be close to the library the next day, the memory-aid application can remind the user to take the books on loan when exiting the house [AshS03].

Life blogging. Context-aware life blogging is like writing your personal diary in an automated way. The mobile application Context Watcher [Koo06] runs on smart phones, automatically connects to available sensors, logs the information, and generates daily summaries about user’s location, activities and moods. Also in this case, the whereabouts diary could provide high-level semantic location descriptions of the identified places. This could improve the life blog by making it more semantically expressive.

5 Related Work

The recent availability of affordable localization mechanism and the recognition of location as a primary source of context information has stimulated a wealth of works addressing topics related to the whereabouts diary. In general, the originality of our work compared to others is to try to give semantic labels to places in a fully automatic way (either by mining the Web, or by classifying the places on the basis of a suitable Bayesian network).

One of the earliest work trying to automatically compose a diary of users' whereabouts is the PEPYS application [New91]. This application uses IR badges and detectors to track user location in an indoor environment. On the basis of such an information, PEPYS compiles a diary of where the user has been and submits it to the users as a memory feedback. This kind of indoor localization systems, as well as its more modern incarnation (e.g., [Hah04]) could naturally complement the proposed GPS diary to deal with indoor settings.

The work described in [HigC05] compares three algorithms to cluster continuous GPS readings to find relevant places. Each of these algorithms could replace the clustering algorithm we described at the beginning of Section 2.2[4]. However all these algorithms are only useful to spot relevant places and identify possible recurrent visit to the sample place. The problem of adding semantic information is completely neglected.

The works in [AshS03, KanW04, MarS00, SchR06] presents similar kind of clustering algorithms.

The problem of adding semantic tags is posed, at least as an open problem, in [Hig03]. Other than clarifying the importance and the need for such a conversion from "positions" to "places", the author illustrates two viable approaches to add semantics. The first approach is based on labeling places on the basis of the activities performed in there. The author proposes using RFID tags to infer users' activities on the basis of the objects being touched (e.g., the user touches a fork and a knife, the system infers he is having dinner). Then, the system uses the activity (e.g., having dinner) to label the place (e.g., restaurant). The second approach involves humans assigning labels to places proactively, and exchanging such labels among users. Neither of the two approaches has been actually realized, and they are mainly left as future work. In any case, once available, they could be well complement and integrate our proposal.

The works described in [PatL03, PatL04] adopt a Bayesian network to infer high-level user behaviors from low-level GPS readings. While their approach is similar to ours, their goal is different. While we try to classify the places, they try to classify user activities and eventually predict where the user will go next, on the basis of his past routes. It is worth to report that some user activities can directly identify the place in which they occur. For example, a sharp step in the user speed can reveal the user started/stopped driving the car. This automatically can be used to label that place as a parking place. Such kind of further information could improve the whereabouts diary.

The work presented in [Flan06] presents an interesting algorithm to share labels to reach a consensus on what a place means to a given group of users. Although, this work is interesting and cloud also be someway integrated with the whereabouts diary, we

[4] It is worth noticing that the surveyed algorithm called BeaconPrint is more reliable and precise than the one we implemented, thus it could also improve the performance of our system.

think it has two main drawbacks. On the one hand, the system proposes random labels like “xyz” to places not allowing for a meaningful semantic description. On the other hand, in a lot of situations, we think it does not make sense to reach a consensus. The same place can mean different things to different people. For example, the bar tender can classify his pub as “work”, while clients will classify it as “pub”.

In the end, we think that a lot of ideas in related work could be used to improve the diary. However, some of the ideas discussed in the paper, like the proposed Bayesian network, are original of our proposal.

6 Conclusions and Future Work

In this paper we presented the *whereabouts diary* – an application, running on a GPS-equipped handheld device – that records the relevant places visited by the user and classifies them semantically in an automatic way. Several useful applications that can be built on the basis of the diary have been also discussed.

There are several directions to improve this work:

1. Much more information about the places could be retrieved from the Web. This could be very useful especially in the cases in which localization is precise enough to return a single or a couple of addresses. Such kind of retrieved information could be a precious source of information to estimate the user profile.
2. Commonsense data could be exploited to effectively discriminate among several candidate places [Liu04]. For example, if a person went to a restaurant at noon, it is very unlikely that will go to another restaurant at 2pm.
3. Other kind of sensing devices and algorithms could be employed to extract more information about the place. Moreover, some GPS clustering techniques that have been used in related works [HigC05] could improve the performance of our implementation.

In the end, accuracy will be the key measure in which the diary will be evaluated. If the diary is wrong, the applications that use it risk being rendered useless. Accordingly, improving the diary accuracy along all the above three directions is a fundamental future work.

Of course, implementing some of the applications discussed in Section 4 will be an important step to evaluate the diary and to get real feedbacks of its usefulness. With this regard, we already started implementing the “places for tourist” application and we plan to test it on a large population of users.

Finally, one important aspect that should be carefully considered is related to the privacy implications involved in the diary usage. This is especially important in consideration of the fact that several applications described in the previous section involve the exchange of diaries among users. With this regard, it would be interesting to investigate the possibility of exchanging hash-based signatures of the diary entries, instead of the actual values to preserve privacy [KosN06].

Acknowledgements

Work supported by the project CASCADAS (IST-027807) funded by the FET Program of the European Commission.

References

[AshS03] D. Ashbrook, T. Starner, "Using GPS to Learn Significant Locations and Predict Movement Across Multiple Users", Personal and Ubiquitous Computing, Volume 7(1):275-286, 2003.

[AxuV06] J. Axup, S. Viller, I. MacColl, R. Cooper, "Lo-Fi Matchmaking: A Study of Social Pairing for Backpackers", International Conference on Ubiquitous Computing, Orange County (CA), USA, 2006

[BenM03] B. Benatallah, Z. Maamar, "Special issue on M-services", IEEE Transactions on Systems, Man, and Cybernetics, Part A, 33(6):665-776, 2003.

[CasR07] G. Castelli, A. Rosi, M. Mamei, F. Zambonelli, A Simple Model and Infrastructure for Context-aware Browsing of the World, International Conference on Pervasive Computing and Communication, IEEE CS Press, White Plains (NY, USA), 2007.

[Fit04] M. Fitzgerald, "Your Digital Wallet", MIT Technology Review, 24 August 2004.

[Flan06] J. Flanagan, "An Unsupervised Learning Paradigm for Peer-to-Peer Labeling and Naming of Locations and Contexts", International Workshop on Location and Context-Awareness, Dublin, Ireland, 2006.

[Hah04] D. Hahnel, W. Burgard, D. Fox, K. Fishkin, M. Philipose, "Mapping and Localization with RFID Technology", IEEE International Conference on Robotics and Automation, New Orleans (LA), USA, 2004.

[Hig03] J. Hightower, "From Position to Place", Workshop on Location-Aware Computing, Seattle (WA), USA, 2003.

[HigC05] J. Hightower, S. Consolvo, A. LaMarca, I. Smith, and J. Hughes, "Learning and Recognizing the Places We Go", International Conference on Ubiquitous Computing, Tokyo, Japan, 2005.

[KanW04] J. Kang, W. Welbourne, B. Stewart, G. Borriello, "Extracting places from traces of locations", ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots, Philadelphia (PA), USA, 2004.

[Koo06] J. Koolwaaij, "Context Watcher", http://portals.telin.nl/contextwatcher/index.aspx

[KosN06] V. Kostakos, E. O'Neill, A. Shahi, "Building Common Ground for Face to Face Interactions by Sharing Mobile Device Context", International Workshop on Location and Context-Awareness, Dublin, Ireland, 2006.

[Liu04] H. Liu, P. Singh, "ConceptNet: a practical commonsense reasoning toolkit", BT Technology Journal 22(4):211-226, Kluwer Academic Publishers, 2004.

[MamN07] M. Mamei, R. Nagpal, Macro Programming through Bayesian Networks: Distributed Inference and Anomaly Detection, International Conference on Pervasive Computing and Communication, IEEE CS Press, White Plains (NY, USA), 2007.

[MarS00] N. Marmasse, C. Schmandt, "Location-aware information delivery with commotion", International Symposium on Handheld and Ubiquitous Computing, Bristol, UK, 2000.

[New91] W. Newman, M. Eldridge, M. Lamming, "PEPYS: Generating Autobiographies by Automatic Tracking", European Conference on Computer Supported Cooperative Work, Amsterdam, Netherlands, 1991.

[PatC06] K. Patel, M. Chen, I. Smith, J. Landay, “Personalizing Routes”, ACM Symposium on User Interface Software and Technology, Montreux, Switzerland, 2006.

[PatL03] D. Patterson, L. Liao, D. Fox, and H. Kautz. “Inferring high-level behavior from low-level sensors”. International Conference on Ubiquitous Computing, Seattle (WA), USA, 2003.

[PatL04] D. Patterson, L. Liao, K. Gajos, M. Collier, N. Livic, K. Olson, S. Wang, D. Fox, H. Kautz, “Opportunity Knocks: a System to Provide Cognitive Assistance with Transportation Services”. International Conference on Ubiquitous Computing, Nottingham, UK, 2004

[PatX06] D. Patterson, X. Ding, N. Noack, Nomatic: “Location By, For, and Of Crowds”, International Workshop on Location and Context-Awareness, Dublin, Ireland, 2006.

[RomH06] G.-C. Roman, R. Handorean, R. Sen, “Tuple Space Coordination Across Space and Time”, International Conference on Coordination Models and Languages, Bologna, Italy, 2006.

[SchR06] F. Schmid, K. Richter, Extracting Places from Location Data Streams, International Workshop on Ubiquitous Geographical Information Services, Munster, Germany, 2006.